DOCUMENT RESUME

ED 284 195 CS 008 895

AUTHOR Baker, Linda; Zimlin, Laurie
TITLE Training Children To Use Two Levels of Standards To Evaluate Their Comprehension: Evidence of Generalization and Maintenance.
PUB DATE Apr 87
NOTE 44p.; Paper presented at the Biennial Meeting of the Society for Research in Child Development (Baltimore, MD, April 23-26, 1987).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Cognitive Processes; Elementary Education; Metacognition; *Reader Text Relationship; Reading Ability; *Reading Comprehension; Reading Difficulties; *Reading Instruction; *Reading Processes; Reading Research; Reading Strategies; Reading Tests; Theory Practice Relationship
IDENTIFIERS *Metacomprehension

ABSTRACT

Noting that children evaluate their reading comprehension more effectively when they are informed about the standards of evaluation they should use, a study examined whether training in the use of certain standards would show generalization from one level of processing to the other, and also maintenance of standard use. The study also examined whether subjects would benefit more from training in "macroprocessing" standards (structural cohesiveness, internal consistency, informational completeness) which require more global text integration than "microprocessing" standards (lexical, propositional cohesiveness, external consistency) which involve more local attention to words and sentences. Subjects, 80 average and above average readers in fourth grade, were assigned to one of three instructional groups: macroprocessing, microprocessing, or no instruction. Subjects in the training groups were given practice applying the standards appropriate to their condition. All groups then read and evaluated passages containing problems requiring the use of all standards. They were tested again two and a half weeks later. Results indicated that the children were more likely to identify problems of the types in which they were instructed, but both trained groups identified more noninstructed problems than did the control group, indicating generalization. The training advantage was also maintained over time. In addition, the macroprocessing group performed better than the microprocessing group on the delayed test, indicating a longer term benefit of global evaluation training. Reading skill affected performance, but did not interact with other factors. (JC)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED284195

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Training Children To Use Two Levels of Standards to Evaluate their Comprehension: Evidence of Generalization and Maintenance

Linda Baker and Laurie Zimlin

University of Maryland Baltimore County

Paper presented at the meeting of the Society for Research in Child Development, Baltimore, April 1987.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Linda Baker

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

CS008895

Abstract

Children evaluate their comprehension more effectively when they are informed about the standards of evaluation they should use. This study sought to determine whether training in the use of certain standards would show generalization and maintenance. Also of interest was whether subjects would benefit more from training in "macroprocessing" standards (structural cohesiveness, internal consistency, informational completeness) which require more global text integration than "microprocessing" standards (lexical, propositional cohesiveness, external consistency) which involve more local attention to words and sentences. Eighty fourth graders, average and above average readers, were assigned to one of three instructional groups: macroprocessing, microprocessing, or no instruction. Subjects in the training groups were given practice applying the standards appropriate to their condition. All groups then read and evaluated passages containing problems requiring the use of all standards. They were tested again about 2 1/2 weeks later. Children were more likely to identify problems of the types they were instructed upon, but both trained groups identified more noninstructed problems than the control group, indicating generalization. Moreover, the training advantage was maintained over time. In addition, the macroprocessing group performed better than the microprocessing group on the delayed test, indicating a longer-term benefit of global evaluation training. Reading skill affected performance, but did not interact with other factors.

Training Children To Use Two Levels of Standards to Evaluate their Comprehension: Evidence of Generalization and Maintenance

A concern of many psychologists and educators interested in the development of reading skill is the extent to which young readers monitor their comprehension. Comprehension monitoring consists of two general components, evaluation and regulation. Evaluation entails keeping track of the success or failure of one's ongoing efforts to understand; regulation entails taking the appropriate steps to deal with comprehension failure once it has been recognized (Baker & Brown, 1984). Although a number of studies have reported that even elementary school children do not spontaneously evaluate or regulate their comprehension (e.g., Markman, 1977, 1979), more recent research has shown impressive competencies, provided care is taken in structuring the task demands (e.g., Baker, 1984a, 1984b).

One limitation of much of the research is that it has been based on a fairly narrow definition of comprehension monitoring. Usually, comprehension monitoring has been operationalized as the ability to find one or two particular types of problems embedded within a text. However, as Baker (1985) has argued, successful detection of different types of problems requires that readers adopt different standards for evaluating their understanding. Studies examining the use of only one standard of evaluation at a time may therefore provide misleading information about children's actual skills.

Accordingly, the present study investigated children's use of six different standards of evaluation described by Baker (1985) as central to effective comprehension.[1] We introduce here a distinction not made explicit in the original presentation: Some standards involve a fairly local evaluation of individual words or phrases whereas others involve a more global evaluation, requiring integration of larger segments of text. We call the first group of standards "microprocessing" standards, as they are important in setting up an internal representation of a text's microstructure, and we call the latter group "macroprocessing" standards, as they are involved in the construction of a cohesive macrostructure (cf. Kintsch & van Dyke, 1979).

Each of the evaluation standards will now be described briefly. The first three are the microprocessing standards and the next three are the macroprocessing standards.

1. The lexical standard involves checking that the meaning of an individual word is understood and as such involves a very local type of evaluation. However, given the important role that vocabulary knowledge plays in comprehension, use of this standard is vital. To test for the use of this standard, researchers have embedded either difficult vocabulary or nonsense words into a text (e.g., Baker 1984a, 1984b; Garner, 1981; Paris & Myers, 1981).

2. The external consistency standard involves checking that the ideas in the text are true or plausible with respect to

what one already knows. This standard plays an important role in critical reading. Its use has been tested by embedding false or implausible information into a text (e.g., Baker, 1984a, 1984b; Markman & Gorin, 1981).

3. The propositional cohesiveness standard involves checking that there is a cohesive relationship among propositions sharing a local context. The importance of propositional cohesiveness to text comprehension has been documented in many studies of anaphoric reference. One way that the propositional cohesiveness standard has been studied is to embed ambiguous referents into a text (e.g., Flavell, Spear, Green & August, 1981).

4. The internal consistency standard involves checking that the ideas expressed in the text are logically consistent with one another. This type of evaluation is also recognized as an important component of critical reading. Internal consistency is the standard that has been studied most frequently in comprehension monitoring research (e.g., Baker, 1984a, 1984b; August, Flavell, & Clift, 1984; Markman, 1979, Garner, 1981, Zabrucky & Ratner, 1986). The usual approach is to present information in one sentence that contradicts information presented in another sentence. In order to notice this type of problem, the reader must attempt to integrate propositions that may not be adjacent in the text.

5. The structural cohesiveness standard involves checking that the ideas in a text or paragraph are thematically

compatible. The importance of structural cohesiveness to comprehension has been well established in studies manipulating text structure or the presence or absence of topic sentences. Use of the structural cohesiveness standard has been tested by embedding a sentence within a paragraph that is irrelevant to the topic (e.g., Harris, Kruithof, Terwogt, & Visser, 1981; Williams, Taylor, & Ganger, 1981).

6. The informational completeness standard involves checking that the text contains all of the information necessary to achieve a specific goal, as in evaluating whether a set of instructions is complete or whether sufficient information has been presented to fully understand a particular concept. Use of the standard has been tested by omitting essential or expected information (e.g., Markman, 1977).

Research has revealed both age-related and ability-related differences in the likelihood that subjects will apply particular standards. But as noted earlier, few studies have examined how well children can apply multiple standards as they read. Baker (1984a) provided some preliminary evidence relevant to this question by examining the effectiveness with which children could apply three different standards (lexical, internal consistency, and external consistency). She found that fourth and sixth graders, good and poor readers alike, were capable of using more than one standard at a time to evaluate their comprehension, and that the effect was stronger if the children were specifically told what standards they should use.

Given the evidence that elementary school children have the ability to evaluate their comprehension but they do not do so spontaneously or routinely, instructional efforts to promote comprehension monitoring are clearly warranted. Several training studies have in fact been implemented within the past few years, all of them showing that children's comprehension monitoring skills can be enhanced through direct instruction (e.g., Elliot-Faust & Pressley, 1986; Miller, 1985; Palincsar & Brown, 1984; Paris, Cross, & Lipson, 1984.) However, none of these studies assessed the effects of providing children with explicit instruction in applying several different evaluation standards. We felt it was important to do so because the greater a child's repertoire of standards, the more effectively will he or she be able to evaluate comprehension.

In the present study, fourth grade students, identified as average and above average readers, were randomly assigned to one of three groups. One group was provided with instruction and practice in the use of the three microprocessing standards; a second group was provided with instruction and practice in the use of the three macroprocessing standards; the third group served as a control condition and received no instruction. After the training sessions, subjects were given two narrative passages containing problems corresponding to each of the six evaluation standards, and they were asked to identify parts of the passages that were hard to understand. Two or three weeks later all subjects were tested again with two different passages in order

to assess the durability of the instruction.

It was expected that subjects in the two training conditions would identify more problems than those in the control group, in line with previous studies showing better performance when subjects are provided with information regarding the standards to be used than when no information is provided (e.g., Baker, 1984a; Markman, 1979; Markman & Gorin, 1981). Of particular interest was whether training in the use of standards at one level of evaluation would generalize to the use of standards at the other level.

We hypothesized that generalization would occur, despite evidence to the contrary provided by Markman and Gorin (1981). Subjects in their study were set to expect to find problems of either internal consistency or external consistency or else they were not given any specific information about problem types. Although subjects identified more problems of the type they had been set to expect, they were no more likely to identify the other type of problem than the control subjects. Note, however, that subjects were simply given two examples of one type of problem; they were not given practice applying the relevant standard on their own nor were they told that the test passages might contain other types of problems, as were our subjects. Thus, there is reason to believe that generalization may occur with more intensive training. This would be a highly desirable outcome given the multiplicity of ways that comprehension can fail and the virtual impossibility of providing students with

instruction in all of them.

On theoretical grounds, it could be argued that training in the use of macroprocessing standards should result in more generalization than training in the use of microprocessing standards. This is so because successful use of macroprocessing standards, which promote consideration of relations among ideas throughout a text, has as a prerequisite attention to individual propositions or sentences. Thus, readers first must evaluate their understanding of the text microstructure before they can evaluate understanding of the macrostructure. Accordingly, subjects instructed in macrostructure evaluation may be more likely to identify microstructure problems en route to the identification of macrostructure problems. Also of interest was whether there would be differential maintenance of standard use over time as a function of instructional condition.

Finally, it was expected that the skilled readers would perform better than the less-skilled readers, in line with previous results. Our primary reason for examining two levels of reading skill was to determine whether the skilled and less-skilled readers would benefit equally from the training. Evidence that the same type of training is effective for heterogeneous groups of readers would imply that ability grouping is unnecessary for classroom instruction in comprehension monitoring.

Method

Subjects

Participants in the study were 80 fourth graders (mean age = 9.81 years, SD = 0.34), 42 females and 38 males. (One additional male subject participated in the first session but was not available for the second test session and so was excluded from all analyses.) Subjects were chosen for participation on the basis of their scores on the reading subtest of the California Achievement Test. Subjects identified as above average readers had stanine scores of 7, 8, and 9. Those identified as average readers had stanine scores of 4, 5, and 6. There were 45 subjects in the above average group (mean stanine score = 7.70, SD = 0.79) and 35 subjects in the average group (mean stanine score = 5.03, SD = 0.91). Subjects were drawn from three suburban elementary schools with varying demographic characteristics.

Subjects within skill levels were randomly assigned to one of the three treatment conditions with the constraint that the distribution of stanine scores be roughly comparable across groups. To the extent possible, the groups were also balanced with respect to gender. Within conditions, subjects were randomly assigned to receive one of two counterbalanced passage sequences across test days. Half of the subjects read two particular test passages on the first test day and the two remaining passages on the second test day. For the remaining subjects, the sequence was reversed. The order of passage

presentation within sessions was also counterbalanced.

Materials

The materials consisted of instructional worksheets and practice passages for subjects in the two training conditions and four test passages. Both the instructional materials and the test passages contained problems that required the use of six different standards of evaluation for their detection, three of which were considered to be microprocessing standards and three macroprocessing. The nature of each problem type will now be described. Table 1 provides a summary of the six different standards and their corresponding problem types and also gives an example of each as it appeared in the instructional materials.

Insert Table 1 about here

The three microprocessing problems were as follows:

1. Lexical problems consisted of nonsense words that varied in length from one to three syllables. The nonsense words were orthographically regular and served as "nouns," "adjectives," or "verbs." The problems were created by replacing one word in the text with a nonsense word. The context was such that subjects could make a reasonable guess as to the meaning of the nonsense word, but a number of alternatives were equally plausible.

2. External consistency problems consisted of facts that were either untrue or that violated everyday prior knowledge. The problems were created by replacing one word in a sentence

with another word that made the information untrue or implausible.

3. Propositional cohesiveness problems consisted of ambiguous anaphoric referents. They were created by substituting a pronoun or a general noun for a specific noun in a context where two or more specific nouns could be the intended referent for the anaphoric expression.

The three types of macroprocessing problems were as follows:

1. Internal consistency problems consisted of information in one sentence that was inconsistent with information presented in a prior sentence. The inconsistencies were explicit contradictions that could be detected without the need for inferences. The inconsistencies were created by replacing one word with another word conveying an opposite or incompatible meaning. There was one sentence intervening between the two contradictory sentences in two of the test passages. and two sentences intervening in the other two test passages.

2. Structural cohesiveness problems consisted of sentences that were unrelated to the other sentences in a paragraph. Although these sentences were semantically linked to the general theme of the passage, they provided information that was irrelevant to the topic of the paragraph. The structural cohesiveness problems were created by adding a new sentence to each passage rather than modifying existing sentences.

3. Informational completeness problems consisted of incomplete information. They were created by omitting

information that the reader was led to expect would be provided. Prior context implied that a particular number of events or items would be described, but only some of the information was in fact presented.

Instructional materials. The instructional materials consisted of separate worksheets for each standard that were identical in format. At the top of the page was a name for the type of problem corresponding to the target standard: unknown words (lexical), unclear nouns and pronouns (propositional cohesiveness), untrue information (external consistency), contradictory information (internal consistency), unrelated sentences (structural cohesiveness) and missing information (informational completeness). Following was a statement that this specific type of problem can make a passage hard to understand. Next on the sheets were three examples illustrating the problem. (See Table 1.) In each example, the specific part of the sentence or parts of multiple sentences that were problematic were underlined. The examples were generally short, but the length varied because different standards require more or less context to set up the problem. The length ranged from one sentence to four sentences. Next on the sheet were three practice exercises, preceded by instructions to read each sentence carefully, underlining any problems that were found. Each practice exercise contained one instance of the target problem. The exercises also varied from one to four sentences in length as necessary.

On a separate worksheet appeared three additional practice exercises for each problem type. These worksheets were only given to those children who had difficulty with the other practice items.

Instructional materials also consisted of six short narratives, each three or four sentences in length. Three of these narratives were typed on a single sheet of paper and were presented to the microprocessing group; the other three were typed on another sheet of paper and were presented to the macroprocessing group. Each passage contained two of the three problem types presented during training, such that a total of two instances of each problem type appeared over the three passages.

Test Passages. The test passages were four narrative stories written by the authors, each of which contained seven paragraphs and was about 300 words long (M = 297; range = 261-332). The vocabulary was selected from the Harris-Jacobsen readability list (Harris & Sipay, 1981) and was at the second grade level. The passages dealt with visits to places familiar to children through personal experience or the media: Washington, D. C.; the beach; an amusement park; and a pet shop. Each passage was modified to include six different problems, each of which could be detected using a different standard of evaluation. The "microprocessing" problems and the "macroprocessing" problems alternated throughout each passage. No problems appeared in either the first or last paragraphs. Table 2 provides a sample passage. Each problem is underlined

and the nature of the standard necessary for its detection is indicated in brackets.

Insert Table 2 about here

The passages were informally tested with several adults in order to ascertain that the problems were indeed readily perceptible. All of the intended problems were identified by the adults. A few additional unintended problems were also noted, and the stories were modified to eliminate these unintended problems.

Procedure

Children were tested individually in a vacant room of their school during the school day by one of two female experimenters. The majority of subjects were seen by one experimenter on the first test day and by the other experimenter on the second test day. All sessions were tape-recorded on audiotape. The experimenters also recorded subjects' oral comments on prepared response sheets.

The instructional sequence for the two training groups was identical; only the materials differed. Children in the microprocessing group were instructed in the use of the lexical standard, the external consistency standard, and the propositional cohesiveness standard. Children in the macroprocessing group were instructed in the use of the internal consistency standard, the structural cohesiveness standard, and

the informational completeness standard. The three target standards were introduced to each subject in a random order to preclude practice or fatigue effects. The experimenter began by describing each standard and giving three examples of problems requiring the use of the standard. The experimenter read aloud each of the examples and then modeled the application of the relevant standard and detection of the problem. Subjects were encouraged to read along silently on their own copy of the materials. Then the subjects were asked to read and evaluate three additional short text segments appearing at the bottom of the sheet. They were instructed to underline each problem they detected using the appropriate standard and subsequently were asked to explain the nature of the problem. Feedback was provided. If the subject had difficulty identifying the problems, three additional practice items were provided. The experimenter then moved on to the next standard. After receiving instruction in the use of each standard individually, subjects were given practice applying more than one standard at a time. They were provided with three short narratives, each of which contained two of the three target problems. Subjects were again asked to underline the problems they detected, and they were asked to explain the nature of the problems. If subjects had difficulty with this part of the task, they were prompted first and then corrective feedback was provided.

Upon completion of the training exercises, subjects were given two test passages to read. They were instructed to look

for problems in the text, underlining whatever problems they found. They were explicitly told that there might be other kinds of problems in the passages than those they had just learned about. After subjects read each passage, they were asked to explain the nature of the problems they identified. The entire session lasted from 20 to 30 minutes.

Subjects receiving no instruction participated only in the test part of the session. They were told that they would be asked to read some stories and point out the problems that they found. A problem was defined as "something that might confuse people or something that people might have trouble understanding." They were also told that there were many kinds of problems. They were asked to read each story carefully, underlining any problems they found, and then explaining the nature of the problems. These sessions lasted no more than 10 minutes.

The original plan was for all follow-up testing to be conducted about two and a half weeks after the initial test session. However, because of scheduling difficulties, the test-retest interval for some subjects was shorter and for others it was longer. Overall, the mean test-retest interval was 17.30 days (SD = 5.7). Across the different training conditions, the mean test-retest intervals were not significantly different (Ms = 17.22, 16.93, and 17.25 for the microprocessing, macroprocessing, and noninstructed groups, respectively). The procedure for the delayed test was the same as for the noninstructed subjects in

the initial session. Subjects were presented with the two passages they had not seen previously and were asked to identify the problems they contained. Delayed test sessions lasted no more than 15 min.

Results

Performance during Training

The majority of the subjects in both the microprocessing and macroprocessing training conditions performed well on the practice exercises. Recall that during the initial instruction on the use of each standard, all subjects received three practice exercises. If subjects were unsuccessful at identifying the intended problems in these exercises even after prompting, additional practice exercises were provided. Only six subjects required the additional practice, three of whom were above-average readers and three of whom were average readers. Three of the subjects required additional practice on only one problem type: two on two types, and one on all three types. Four of the subjects had difficulty identifying unclear pronouns, two each had difficulty identifying internal inconsistencies and unrelated sentences, and one had difficulty with unknown words. Only the subject who required additional practice on all three problems was still unsuccessful after the additional practice. Thus, it is clear that subjects were capable of applying the appropriate standards when specifically instructed to do so under conditions designed to maximize their performance.

Recall that upon completion of instruction in each standard

individually, subjects were given three short passages to read and were instructed to use all three trained standards to find the embedded problems. The majority of the subjects did well on this task. In fact, a total of 20 out of 55 subjects identified all 6 problems (two requiring the use of each trained standard) completely independently, and an additional 14 subjects identified all problems with minimal prompting. Overall, the mean number of problems independently identified was 4.62 out of 6 possible (SD =1.51). The data were subjected to a 2 (condition) by 2 (skill level) between-subjects analysis of variance. The analysis revealed that neither the main effects nor the interaction were reliable, Fs < 1 . This suggests that the training was equally successful in both conditions. In addition, it indicates that both skilled and less-skilled readers were able to apply the trained standards in simple passages.

Performance on the Test Passages

Subjects' responses were scored for the number of problems of each type that were correctly identified. A problem was only scored as identified if subjects underlined it and gave an appropriate justification as to why they underlined it. (Although it did not happen frequently, there were occasions when target information was identified as problematic for idiosyncratic reasons.) All of the protocols were scored by one of the experimenters, and 50% of them were independently scored by the second experimenter. Overall agreement was 95%, and discrepancies were resolved through discussion.

Effects of training on the use of microprocessing and macroprocessing standards. The dependent measure of primary interest was the number of problems identified at each level of evaluation. Because we were not interested in comparisons among problem types, we collapsed across the three problem types at each level to obtain a measure providing six observations per test session rather than two. Two separate analyses of variance were conducted, one examining performance on the first test day only and the second comparing performance across test days. The analyses consisted of two 3 (instructional condition: microprocessing, macroprocessing, none) by 2 (reading skill level: average, above average) by 2 (passage set order) by 2 (standard level: microprocessing, macroprocessing) [by 2 (test time: immediate, delayed)] mixed factorials. The first three factors were between-subjects and the last [two] within-subjects. (A preliminary analysis including gender as a factor revealed neither a main effect of gender nor any interactions ($F < 1$); for ease of exposition, gender was therefore dropped as a factor.) Tukey's HSD procedure was used for all post-hoc comparisons, $p < .05$.

Consider first performance on the initial test day only. As expected, there was a reliable main effect of instructional condition, $F(2, 68) = 25.26$, $p < .001$. Subjects receiving instruction in the use of either microprocessing or macroprocessing standards identified more problems than subjects receiving no instruction (Ms = 3.25, 3.21, and 1.25, for the

three groups respectively.) Also as expected, there was a reliable interaction between instructional condition and standard level, $F(2, 68) = 26.65$, $p < .001$. Figure 1 illustrates this interaction. Subjects receiving instruction in the use of microprocessing standards identified more microprocessing problems than macroprocessing problems. Similarly, subjects receiving instruction in the use of macroprocessing standards identified more macroprocessing problems than microprocessing problems. Subjects receiving no instruction performed comparably on both levels. Note, however, that both groups receiving instruction outperformed the control group even on the noninstructed problems. This provides clear evidence that subjects generalized their training from one level of standard application to the other. Contrary to expectations, the macroprocessing group did not exhibit more generalization than the microprocessing group.

Insert Figure 1 about here

As expected, better readers identified more problems than average readers, $F(1, 68) = 12.18$, $p < .001$. The means were 2.98 and 2.15 for the two groups, respectively. However, there was no evidence that reading skill had differential effects on training outcomes; none of the interactions involving reading level were reliable, all $ps > .10$. In other words, both groups benefitted equally from the training and showed comparable levels of

generalization.

The only other reliable effect was a main effect of passage set, $F(1, 68) = 17.62$, $p < .001$. Subjects reading the two passages in one set identified more problems overall than subjects reading the two passages in the other set (Ms = 3.10 and 2.00). However, because passage set did not interact with other factors, the main effect does not compromise interpretation of the effects of interest.

Consider now the results of the second analysis which included time of test as a factor. The main effects of instructional condition and reading skill level were again reliable, $F(2, 68) = 24.15$, $p < .001$, and $F(1, 68) = 16.32$, $p < .001$, respectively. The trained groups continued to outperform the untrained group, and above average readers outperformed the average readers. The main effect of test time was not reliable, $p > .10$, but the two-way interactions with instructional condition and with standard level were reliable, $F(2, 68) = 27.43$, $p < .001$, and $F(1, 68) = 12.58$, $p < .001$, respectively. In addition, the three-way interaction of test time by instructional condition by standard level was reliable, $F(2, 68) = 3.14$, $p < .05$. The three-way interaction is illustrated in Figure 2. The figure shows that the interaction is attributable to the fact that the performance of the subjects receiving microprocessing instruction declined on the delayed test for microprocessing problems (but not for macroprocessing problems). In contrast, there were no declines in performance for either the

macroprocessing instruction group or for the no instruction group from the immediate test to the delayed test. Thus, it appears that the training for the macroprocessing group was somewhat more durable than the training for the microprocessing group. Nevertheless, the microprocessing group continued to identify more microprocessing problems than macroprocessing problems and they continued to identify more problems overall than the noninstructed subjects, indicating that they still retained some of the benefits of training in evaluating their understanding.

Insert Figure 2 about here

The only other reliable effect was an interaction of passage set with standard level, F 1, 68) = 41.73, $p < .001$. Recall that half of the subjects read two particular passages on the first test day and two other passages on the second test day; for the remaining subjects the pattern was reversed. It appears that subjects who read one set of passages on the first day performed better overall on the macroprocessing problems than the microprocessing problems, whereas there was no difference in performance across the two levels for the other group of subjects. The fact that there was no interaction with test time, however, indicates that the performance advantage was not specific to the two passages in question. If it had been, one group of subjects should have performed relatively better on the macroprocessing problems on the first test day and the second

group on the second test day. Thus, it is difficult to account for the interaction involving passage set, but, again, it does not seem to compromise interpretation of the effects of interest.

Effects of training on the application of individual standards. Although our primary interest was in determining whether subjects showed general benefits of instruction in evaluating their understanding, we also sought to determine whether the training had differential effects depending on the specific standard required. Accordingly, a second set of analyses were conducted where the dependent variables were the number of problems identified using each standard. Each analysis consisted of a 3 (instructional condition) by 2 (reading level) by 2 (time of test) mixed factorial.

Figures 3 and 4 present the mean number of problems identified for the microprocessing and macroprocessing standards, respectively, collapsing across time of test and reading level. As can be seen, subjects who had received training in the use of each specific standard were consistently superior to subjects who had received no instruction at all. However, the magnitude of the differences between the two instructional conditions varied, as did the magnitude of the differences between noninstructed subjects and those who had received instruction at the alternative level of evaluation. Performance patterns for each of the six standards will be summarized below; only statistically significant results will be noted, except in cases where the lack of reliability is a departure from the usual pattern. Tukey's

HSD procedure was again used for comparisons among means.

Insert Figures 3 and 4 about here

Consider first performance for the three microprocessing standards, shown in Figure 3. For the lexical standard, subjects receiving microprocessing training identified significantly more problems than subjects who received macroprocessing training or no training, and the latter groups did not differ between themselves, $F(2, 74) = 13.51$, $p < .001$. Clearly, there was no generalization of macroprocessing training to the identification of unknown words. The main effect of test time was also reliable, $F(1,74) = 8.37$, $p < .01$, with subjects identifying more problems on the immediate test than on the delayed test. In contrast to the general pattern, skilled readers and less-skilled readers did not differ in their identification of word-level problems, $p > .10$.

For the external consistency standard, subjects receiving microprocessing instruction and subjects receiving macroprocessing instruction identified comparable numbers of problems; both groups differed significantly from the noninstructed subjects, $F(2, 74) = 19.55$, $p < .001$. Thus there was clear generalization of training in macrostructure evaluation to the identification of prior knowledge violations. The effect of reading level was reliable, $F(1, 74) = 8.36$, $p < .01$, but it was qualified by two-way interactions with instructional

condition, $F(2, 74) = 4.48$, $p < .01$, and with test time, $F(1,74) = 5.21$, $p < .01$. The first interaction is attributable to the fact that the better readers identified more problems than the average readers only in the microprocessing and non-instructed conditions; subjects in the macroprocessing condition performed comparably at both skill levels. The test time effect is attributable to the fact that the average readers showed a decline in performance from the immediate test to the delayed test, but the above-average readers did not.

For the propositional cohesiveness standard, subjects receiving microprocessing instruction identified significantly more problems than non-instructed subjects, but the subjects receiving macroprocessing instruction were not significantly different from either group, $F(2, 74) = 15.34$, $p < .001$. In addition, the better readers identified more problems than the less-skilled readers, $F(1, 74) = 15.92$, $p < .001$.

We turn now to consideration of performance on the three macroprocessing standards, shown in Figure 4. For the structural cohesiveness standard, subjects receiving macroprocessing instruction identified significantly more problems than subjects receiving no instruction, but the subjects receiving microprocessing instruction did not differ significantly from either group, $F(2, 74) = 8.06$, $p < .001$. The effect of reading level was but marginally reliable, $p < .10$.

For the internal consistency standard, subjects receiving macroprocessing instruction identified more problems than non-

instructed subjects, with performance of the subjects receiving microprocessing instruction again intermediate to the other two groups, $F(2, 74) = 17.88$, $p < .001$. In addition, skilled readers identified more problems than less skilled, $F(1, 74) = 11.88$, $p < .001$.

Finally, for the informational completeness standard, subjects receiving macroprocessing instruction identified more problems than subjects receiving microprocessing instruction, who in turn identified more problems than subjects receiving no instruction, $F(2, 74) = 19.28$, $p < .001$. Clearly, subjects generalized their microprocessing training to the identification of missing information. Again, skilled readers identified more problems than less-skilled readers, $F(1, 74) = 7.48$, $p < .01$.

Discussion

The major contribution of this study is its demonstration of generalization and maintenance when students were trained to evaluate their comprehension using several different standards of evaluation. The finding that trained students were more proficient at evaluating their comprehension according to both trained and nontrained criteria is particularly valuable from an instructional standpoint because it would be difficult if not impossible to instruct students on every possible type of comprehension problem they might encounter.

The fact that macroprocessing instruction proved to be more long-lasting suggests that we should focus our instructional efforts on teaching students to consider the relations among

ideas from different parts of a text. Such instruction would also have the effect of fostering students' ability to construct a coherent representation of a text, a crucial goal of comprehension. Although the macroprocessing instruction did not have the hypothesized effect of stimulating relatively more evaluation at the microprocessing level as well, it does seem that subjects were better able to retain and/or implement the specific knowledge they had acquired. What would undoubtedly be most beneficial, of course, would be instruction in using standards at both levels of evaluation. This would certainly be feasible, especially given the ease with which the students in both instructional groups handled the training exercises. Indeed, research recently completed has shown that third and fifth graders show clear benefits of instruction in the use of all six standards, even when the instruction is conducted within a single 30-minute session.

It should be noted that even though subjects benefitted from instruction, the overall levels of problem detection were still quite low. In fact, the average performance of the instructed groups was only 54% on the test passages, even though their performance on the training exercises that specified which standard to use was above 90%. It appears that the transition from evaluation of short segments of text with specified standards to the evaluation of lengthy passages requiring multiple standards is not mastered rapidly. The fact that performance was at intermediate levels (77%) on the short

training passages where subjects needed to use all three trained standards supports this observation. To effect full mastery, students would need considerably more practice and feedback than they received in the present study, and they would need to be provided with increasingly complex materials.

The finding that above-average readers identified more problems overall than average readers was not surprising, and is consistent with virtually all comprehension monitoring studies that have included comparisons of skilled and less-skilled readers. What was encouraging, however, was that skill level did not interact with training condition, level of standard, or test time. In other words, all readers benefitted from training in both macroprocessing and microprocessing evaluation; all showed comparable levels of generalization; and all showed comparable maintenance over time. It is not clear, of course, whether similar results would obtain if we had included a sample of truly poor readers. However, in the school system in which we worked, reading scores in the fourth stanine are considered to be quite low, even though they are considered average on a nationwide basis. Thus, our sample of "average" readers did include a majority who were not achieving as well as their peers, and so our results do have implications for the improvement of comprehension monitoring skills among lower-achieving students.

Although overall levels of performance were equally good for both the microprocessing and macroprocessing standards, there were some variations in performance across individual standards.

These variations are to be expected given the differences in the cognitive processes needed for their application, as well as the differences in the textual materials in which the problems were embedded. Thus, comparisons across standards were of less theoretical interest than comparisons within standards. We were particularly interested in learning that the advantage of specific instruction over no instruction held for each of the six standards. There were, however, differential levels of generalization of training. Some of these effects will be discussed below.

The fact that there was absolutely no generalization of macroprocessing instruction to the use of the lexical standard is consistent with previous research showing that students frequently do not identify word-level problems unless they are specifically instructed to do so (e.g., Baker, 1984b). However, it may not be that students fail to use the lexical standard, but rather that they are reluctant to acknowledge that they do not know the meanings of words that are short and readily decodable. Evidence in support of this suggestion was provided by Baker (1987), who found a consistent reporting advantage for longer nonwords over shorter nonwords.

That the macroprocessing condition identified as many external consistency problems as the microprocessing condition can also be interpreted in accordance with previous research. Children have a tendency to question the external consistency of a text any time they are set to evaluate it critically (e.g.,

Baker 1984a, 1984b; Markman, 1979). Thus, the external consistency standard is in some ways a default standard for young readers. (The lexical standard is also a default standard, but one which is harder to show empirically because of the response bias problem.) The finding is inconsistent, however, with Markman and Gorin (1981), who found no generalization of instruction to use the internal consistency standard (a component of our macroprocessing instruction) to the use of the external consistency standard (a component of our microprocessing instruction). However, as noted in the introduction, there were substantial differences between the studies which may account for the differences in results.

The informational completeness standard was the only other standard for which generalization of training reached conventional significance levels, although generalization for the internal consistency and propositional cohesiveness standards was also fairly strong. There was less generalization for the structural cohesiveness standard, but overall performance levels were quite low. In sum, performance on the individual standards generally mirrored the overall effects, illustrating that the beneficial effects of instruction are not standard-specific.

In conclusion, the present research has shown that students can learn to apply several different standards to evaluate their comprehension. The next step is to ascertain whether more proficient use of the various standards results in improvements in comprehension per se. The fact that each of the standards

examined in the present study plays a crucial role in comprehension (cf., Baker, 1985) suggests that improvements in comprehension are highly likely.

References

August, D. L., Flavell, J.H., & Clift, R. (1984). Comparison of comprehension monitoring of skilled and less skilled readers. Reading Research Quarterly, 20, 39-53.

Baker, L. (1984). Spontaneous versus instructed use of multiple standards for evaluating comprehension: Effects of age, reading proficiency and type of standard. Journal of Experimental Child Psychology, 38, 289-311. (a)

Baker, L. (1984) Children's effective use of multiple standards for evaluating their comprehension. Journal of Educational Psychology, 76, 588-597. (b)

Baker, L. (1985). How do we know when we don't understand? Standards for evaluating text comprehension. In D. L. Forrest-Pressley, G. E. MacKinnon, & T. G. Waller (Eds.), Metacognition, cognition, and human performance (pp. 155-205). NY: Academic Press.

Baker, L. (1987). Developmental change in readers' responses to unknown words. Unpublished manuscript, University of Maryland Baltimore County.

Baker, L., & Brown, A. L. (1984). Metacognitive skills and reading. In P. D. Pearson (Ed.), Handbook of research in reading (pp. 353-393). NY: Longman.

Elliott-Faust, D. J., & Pressley, M. (1986). How to teach comparison processing to increase children's short-and long-term listening comprehension monitoring. Journal of Educational Psychology, 78, 27-33.

Flavell, J. H., Speer, J. R., Green, F. L., & August, D. L. (1981). The development of comprehension monitoring and knowledge about communication. Monographs of the Society for Research in Child Development, 46, (Whole No. 192).

Garner, R. (1981). Monitoring of passage inconsistency among poor comprehenders: A preliminary test of the "piecemeal processing" explanation. Journal of Educational Research, 74, 159-162.

Harris, P. L., Kruithof, A., Terwogt, M. M., & Visser, P. (1981). Journal of Experimental Child Psychology, 31, 212-230.

Kintsch, W., & Van Dijk, T. A. (1978). Toward a model of text comprehension and production. Psychological Review, 85, 363-394.

Markman, E. M. (1977). Realizing that you don't understand: A preliminary investigation. Child Development, 46, 986-992.

Markman, E. M. (1979). Realizing that you don't understand: Elementary school children's awareness of inconsistencies. Child Development, 50, 643-655.

Markman, E. M., & Gorin, L. (1981). Children's ability to adjust their standards for evaluating comprehension. Journal of Educational Psychology, 73, 320-325.

Miller, G. (1985). The effects of general and specific self-instruction training on children's comprehension monitoring performance during reading. Reading Research Quarterly, 20, 616-628.

Palincsar, A. S., & Brown, A. L. (1984). Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities. Cognition and Instruction, 1, 117-175.

Paris, S. G., Cross, D. R., & Lipson, M. Y. (1984). Informed strategies for learning: A program to improve children's reading awareness and comprehension. Journal of Educational Psychology, 76, 1239-1252.

Paris, S. G., & Myers, M. (1981). Comprehension monitoring, memory, and study strategies of good and poor readers. Journal of Reading Behavior, 13, 5-22.

Williams, J. P., Taylor, M. B., & Ganger, S. (1981). Text variations at the level of the individual sentence and the comprehension of simple expository paragraphs. Journal of Educational Psychology, 73, 851-865.

Zabrucky, K., & Ratner, H. H. (1986). Children's comprehension monitoring and recall of inconsistent stories. Child Development, 57, 1401-1418.

Footnotes

1. Baker (1985) also identified a seventh standard of evaluation, the syntactic standard, which involves evaluation of the grammaticality of a sentence or phrase. Use of this standard has been studied by scrambling the words in a sentence or a part of a sentence (e.g., Paris & Myers, 1981). However, because the resulting syntactic violations also affect the semantic aspects of the sentence, it is difficult to determine whether syntactic evaluation per se triggers awareness of comprehenson difficulty. Accordingly, this standard was not included in the present investigation.

Table 1

Examples of the Problems Corresponding to Each Standard

Microprocessing Standards

1. Standard: Lexical

   Problem: Unknown words (nonsense words)

   Example: Before the frick could yell, the truck drove away.

2. Standard: External consistency

   Problem: Untrue or implausible information

   Example: George cut the paper with a hammer.

3. Standard: Propositional cohesiveness

   Problem: Ambiguous referents

   Example: Dan and Peter were fighting. He got punished.

Macroprocessing Standards

1. Standard: Structural cohesiveness

   Problem: Unrelated information

   Example: Shelly's favorite color is blue. She wears blue every day. The sky is blue. Shelly has blue pants, blue shirts, and even blue shoes.

2. Standard: Internal consistency

   Problem: Contradictory information

   Example: The morning light was very bright. Chuck and Kenny wanted to go outside and play. Their mother said no because it was too dark.

3. Standard: Informational Completeness

   Problem: Omitted information

   Example: Jill got five presents for her birthday. Her parents gave her a doll. Her grandparents gave her a dress. Jill's brother gave her a bear.

Table 2

Sample Passage Illustrating the Six Different Problems and their Corresponding Standards

The School Trip

Every year the fourth grade class of Seabrook Elementary School takes a trip in the spring. This time they decided to go to Washington, D.C.

The children had chosen Tuesday, April 15, 1986 for the trip. It was a warm and sunny morning. All of the students were very excited about going to Washington, D.C. It was very foggy outside. [internal consistency problem]

Each child was to pick one partner to sit next to on the bus. Charlie, Greg and Scott were all friends. It was hard to decide which two boys would be partners. Finally, Charlie picked the boy. [propositional cohesiveness problem]

When they arrived in Washington, D.C., the students were told that they would be going to see four places. First they would see the White House and then they would see the Washington Monument. [informational completeness problem]

The first place the school children went to see was the White House. This is where the king and his wife live. [external consistency problem] It is a very big and beautiful house. Each room is a different color. There is a red sofa and a glass table in the red room. There is a blue braff and blue floor tiles in the blue room. [lexical problem] In the green room, the walls are painted green and there is a piano.

After visiting the White House the fourth graders saw the Washington Monument. It is very big. In Washington, everybody takes the subway to work. [structural cohesiveness problem] The monument has many steps. Each child climbed to the top of the building. From the top, one can look out and see the entire city. After climbing down the steps the children were tired and decided to rest.

The children ate lunch on the grass. It was nice. After lunch the children returned to school. When they went home the students told their families about the fun day they had in Washington, D.C.

FIGURE 1

MEAN NUMBER OF PROBLEMS IDENTIFIED AT TWO LEVELS OF EVALUATION AS A FUNCTION OF INSTRUCTIONAL CONDITION ON THE IMMEDIATE TEST

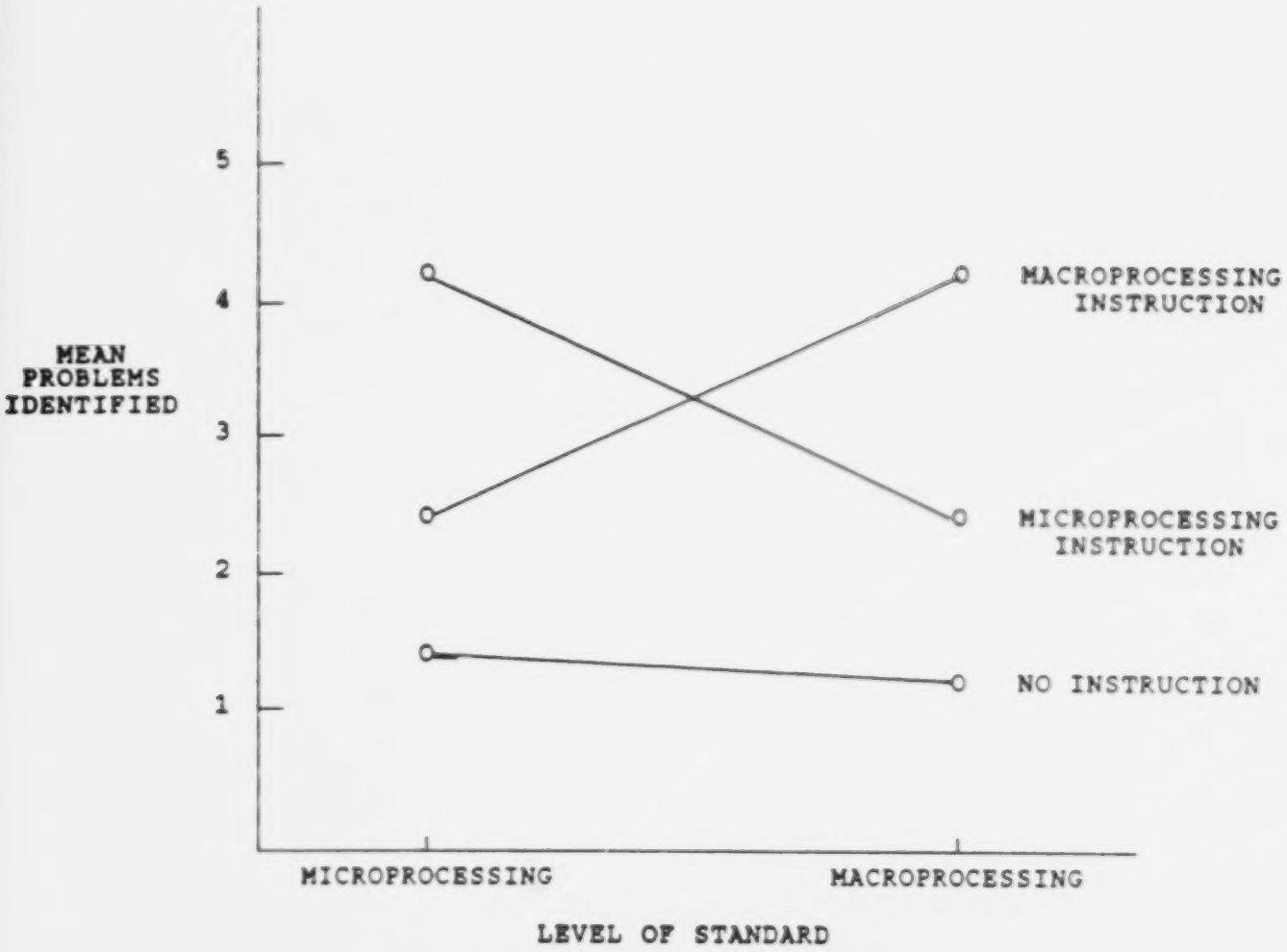

FIGURE 2

MEAN NUMBER OF PROBLEMS IDENTIFIED AT TWO LEVELS OF EVALUATION AS A FUNCTION OF INSTRUCTIONAL CONDITION AND TIME OF TEST

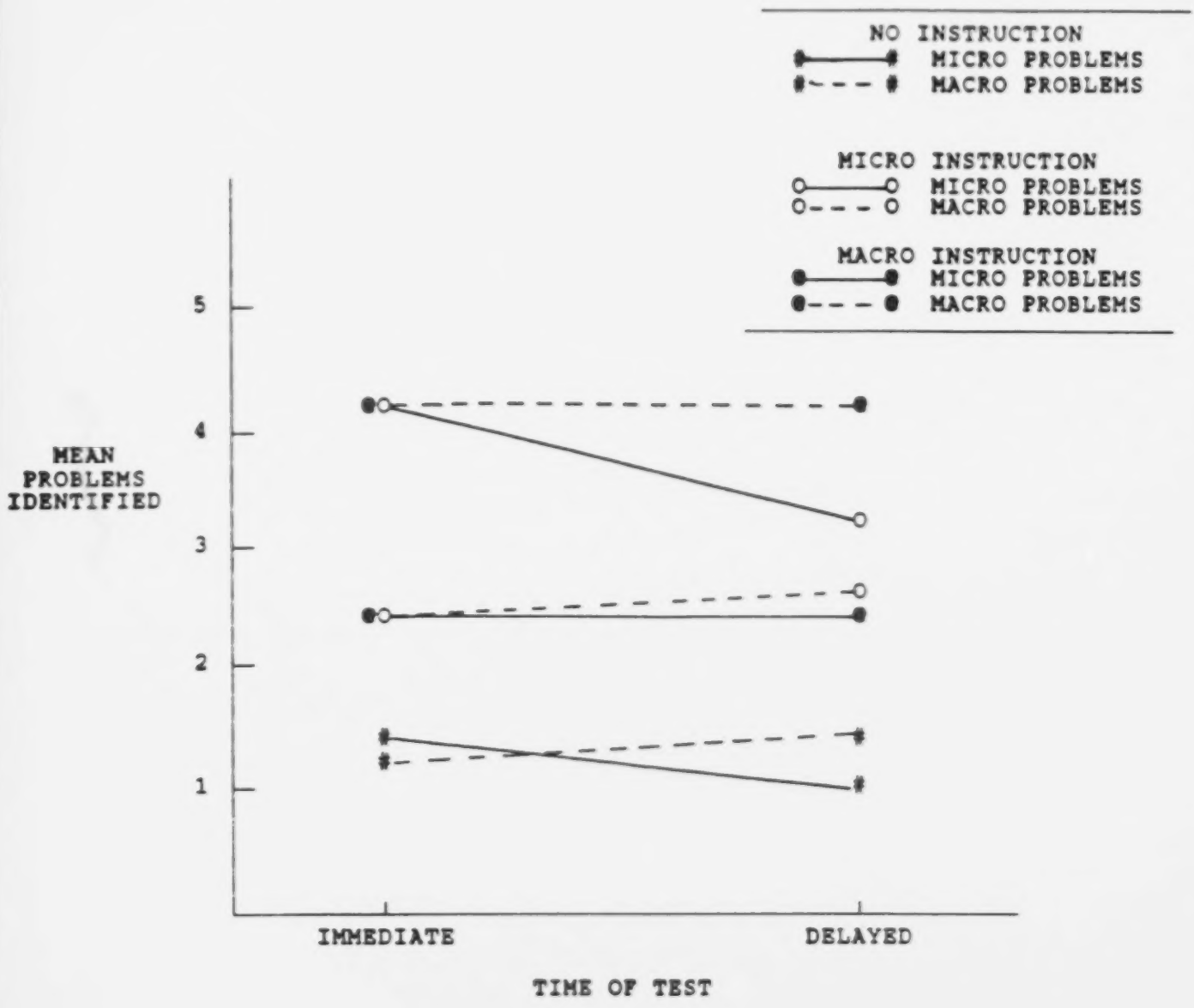

FIGURE 3

MEAN NUMBER OF PROBLEMS IDENTIFIED USING EACH STANDARD

AT THE MICROPROCESSING LEVEL AS A FUNCTION OF INSTRUCTIONAL CONDITION

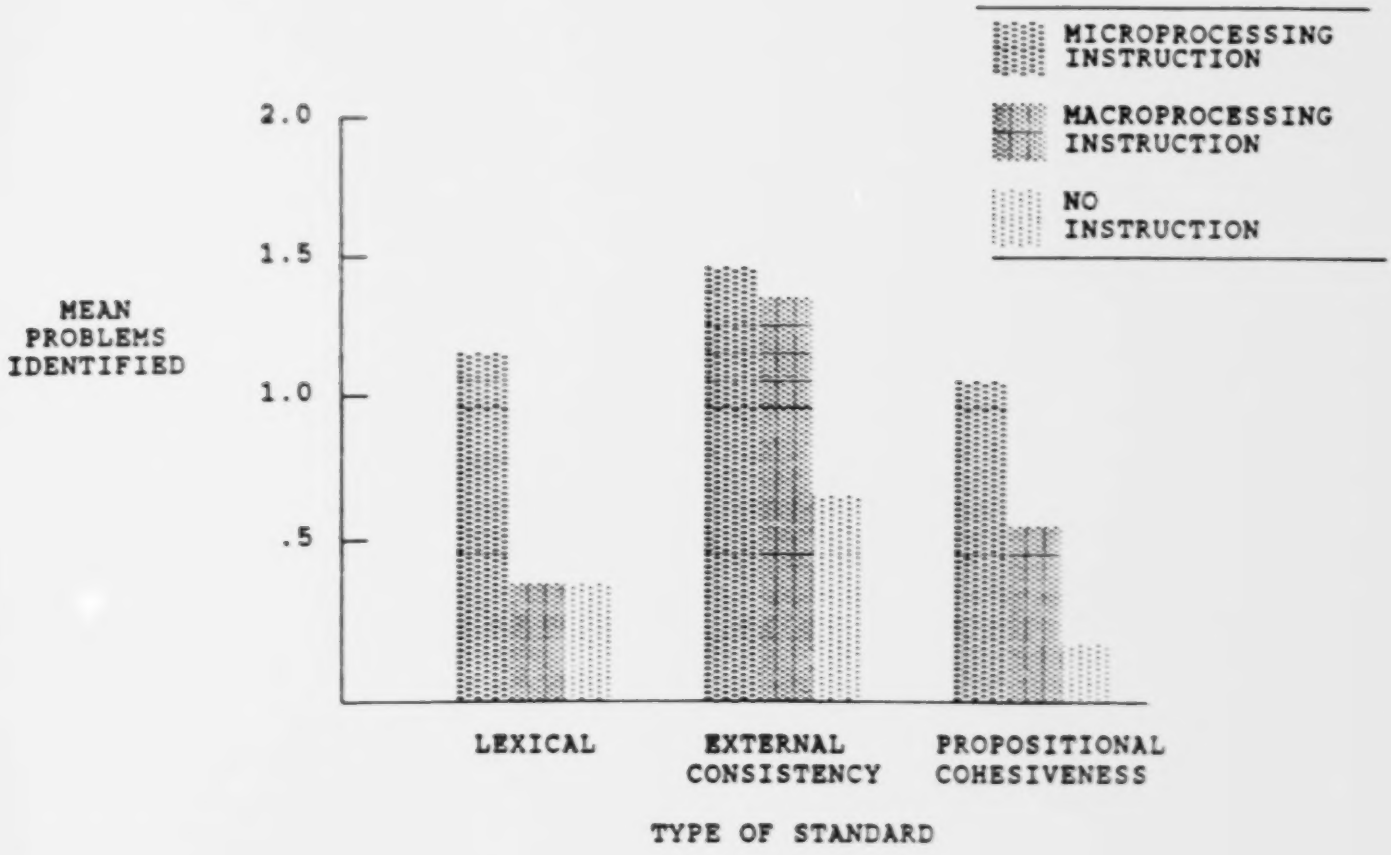

FIGURE 4

MEAN NUMBER OF PROBLEMS IDENTIFIED USING EACH STANDARD

AT THE MACROPROCESSING LEVEL AS A FUNCTION OF INSTRUCTIONAL CONDITION